Post-Pleistocene radiation of the pea aphid complex revealed by rapidly evolving endosymbionts

Jean Peccoud[a], Jean-Christophe Simon[a], Heather J. McLaughlin[b], and Nancy A. Moran[b,1]

[a]Unité Mixte de Recherche 1099 Biologie des Organismes et des Populations appliquée à la Protection des Plantes, Institut National de la Recherche Agronomique, Domaine de la Motte BP 35327, 35653 Le Rheu cedex, France; and [b]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85718

Edited by David M. Hillis, University of Texas, Austin, TX, and approved August 5, 2009 (received for review May 9, 2009)

Adaptation to different resources has the potential to cause rapid species diversification, but few studies have been able to quantify the time scale of recent adaptive radiations. The pea aphid, *Acyrthosiphon pisum,* a model of speciation for host-specialized parasites, consists of several biotypes (races or species) living on distinct legume hosts. To document this radiation, we used rapidly evolving sequences from *Buchnera*, the maternally transmitted bacterial endosymbiont of aphids. Analyses of *Buchnera* pseudogene sequences revealed that 11 host-associated biotypes sort mostly into distinct matrilines despite low sequence divergence. A calibration based on divergence times of 7 sequenced genomes of *Buchnera* allowed us to date the last maternal ancestor of these biotypes between 8,000 and 16,000 years, with a burst of diversification at an estimated 3,600–9,500 years. The recency of this diversification, which is supported by microsatellite data, implies that the pea aphid complex ranks among the most rapid adaptive radiations yet documented. This diversification coincides with post-Pleistocene warming and with the domestication and anthropogenic range expansion of several of the legume hosts of pea aphids. Thus, we hypothesize that the new availability or abundance of resources triggered a cascade of divergence events in this newly formed complex.

adaptive radiation | *Buchnera* | ecological speciation | host races | Neolithic agriculture

The evolution of ecological diversity within a rapidly multiplying lineage constitutes an adaptive radiation (1, 2), a process that may have substantially contributed to Earth's biodiversity (3). Documented cases of recent adaptive radiation illustrate how divergent selection on ecologically relevant traits promotes reproductive isolation and repeated speciation events (reviewed in refs. 4 and 5). In the young Lake Victoria, cichlid fish repeatedly adapted to living at different depths through their sensitivity to specific light spectra and different male nuptial colors, leading to sexual isolation and ultimately speciation (6). Theoretical work suggests that this process of ecological speciation can generate numerous species within a few thousand generations, in a burst of diversification following the invasion of a rich and diverse environment (7). Identification of ongoing adaptive radiations, combined with estimates of accompanying diversification rates, could validate theoretical predictions, but these measures remain rare.

Phytophagous insects frequently present closely related species or races (i.e., populations in partial reproductive isolation) associated with distinct host plant species (8, 9), which potentially constitute recent adaptive diversifications (10, 11). Because the host plant is often the location for mating, reproduction, and development, the shift to a novel host may directly result in premating isolation (12, 13) and ecologically unfit hybrids (14, 15), favoring the formation of new entities.

The pea aphid, *Acyrthosiphon pisum* Harris, is a long-standing model system for the initial stages of speciation through host race formation (16, 17). In their Palearctic native range, pea aphid populations are distinguished by their specialization to specific legume hosts (16, 18–21). Nuclear DNA markers indicate the presence of at least 4 cryptic species, 1 of which comprises 8 host races showing partial reproductive isolation (22) (Table 1); we refer to these collectively as "biotypes." Thus, the pea aphid complex appears to represent early stages of an adaptive radiation, as supported by very low polymorphism of both nuclear and mitochondrial DNA sequences (22, 23). This low sequence variation and the unknown mutation rates for these genetic markers have hindered estimation of the age of the pea aphid complex, however. *Buchnera aphidicola*, the obligate bacterial endosymbiont of aphids, shows strict maternal inheritance and thus phylogenetic relationships that track host matrilines (female lineages) (24–26). *Buchnera* of pea aphids (hereinafter called *Buchnera*-Ap) also undergoes a very high mutation rate, which enables dating of recent divergences (27). Here we used 3 rapidly evolving pseudogene regions and 7 previously sequenced genomes of *Buchnera*-Ap to date the radiation of the pea aphid complex.

Results

Most Pea Aphid Biotypes Are Associated with Distinct *Buchnera*-Ap Haplotypes. Fig. 1 shows the haplotype network based on *Buchnera*-Ap sequenced for 2 regions (alignment size, 1.5 kb) for 365 individuals collected from 24 host plants worldwide [see supporting information (SI) Table S1 for collection information and GenBank accession numbers]. Most come from localities in or near the probable native range of the pea aphid (Europe, Near East, and North Africa), with a minority from introduced populations (Americas, East Asia, and Australia). Very low homoplasy is observed within *Buchnera*-Ap, as expected for clonal lineages showing low divergence. Almost all samples fall within a tight cluster (clade 1, Fig. 1) with maximum divergence of 0.6%. Three collections from Japan are clustered separately as a single, more divergent lineage.

Of the 365 samples, 254 from Europe and Chile were previously assigned to host-specialized biotypes (colored symbols in Fig. 1) based on nuclear microsatellite markers and host specialization tests (22, 28). With the exception of the races associated with *Medicago sativa* and *M. lupulina*, whose matrilines are scattered almost through the network (in red and brown, respectively), individuals assigned to a given biotype bear the same or a closely related *Buchnera* haplotype. This result supports restricted maternal gene flow between most biotypes. In partic-

Author contributions: J.P., J.-C.S., and N.A.M. designed research; J.P., J.-C.S., H.J.M., and N.A.M. performed research; J.P., J.-C.S., and N.A.M. contributed new reagents/analytic tools; J.P., H.J.M., and N.A.M. analyzed data; and J.P., J.-C.S., and N.A.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. FJ357501–FJ357537, FJ773396–FJ773991, FJ869424–FJ869781, and FJ948190–FJ948191).

[1]To whom correspondence should be addressed. E-mail: nmoran@email.arizona.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0905129106/DCSupplemental.

Table 1. Host range and taxonomic status of 11 biotypes of the pea aphid

Host range	Status*
Cytisus scoparius	Species
Lathyrus pratensis	Species
Ononis repens, O. spinosa	Species
Pisum sativum, annual *Vicia*, *Lens culinaris*	Host race
Vicia cracca	Host race
Medicago sativa	Host race
Medicago lupulina	Host race
Trifolium spp	Host race
Lotus corniculatus	Host race
Lotus pedunculatus	Host race
Melilotus albus, M. officinalis	Host race

*Species status is assigned if no hybridization has been detected in the field; otherwise, host race status is assigned (22).

ular, the cryptic species on *Lathyrus pratensis* (orange) and the race on *Lotus pendunculatus* (light blue) form monophyletic clades. Conversely, the matrilines associated with *Cytisus scoparius* (gray) and *Melilotus* species (dark green) are poorly separated. Many of the samples not assigned to biotypes fall within the cluster corresponding to their collection plant. In particular, most *Trifolium*-collected individuals in introduced populations in Japan, and some in North America, have haplotypes characteristic of the *Trifolium* host race in Europe.

Besides this host-associated clustering, some geographic clustering is apparent for several samples for which host specialization was not characterized, including those from East Asia, Iran, and Tunisia (Fig. 1).

Recent Diversification of *Buchnera*-Ap Matrilines. A previous study estimated the maximum divergence time of 7 fully sequenced genomes of *Buchnera*-Ap at 8,300–15,800 years ago, based on 2 calibration points that support similar rates of sequence evolution (27). The matrilines represented by these 7 genomes coalesce at the base of the main cluster identified here (clade 1 in Fig. 1), that is, at the shared ancestor for all samples except the divergent Japanese lineage. This date provides a good calibration point for dating other divergences of pea aphid matrilines. For this purpose, we increased precision by sequencing a *Buchnera* intergenic region (*prfC-yhgI* spacer) from 175 individuals representative of most of the lineages characterized at the other 2 regions.

Fig. 2 shows the tree corresponding to these samples with

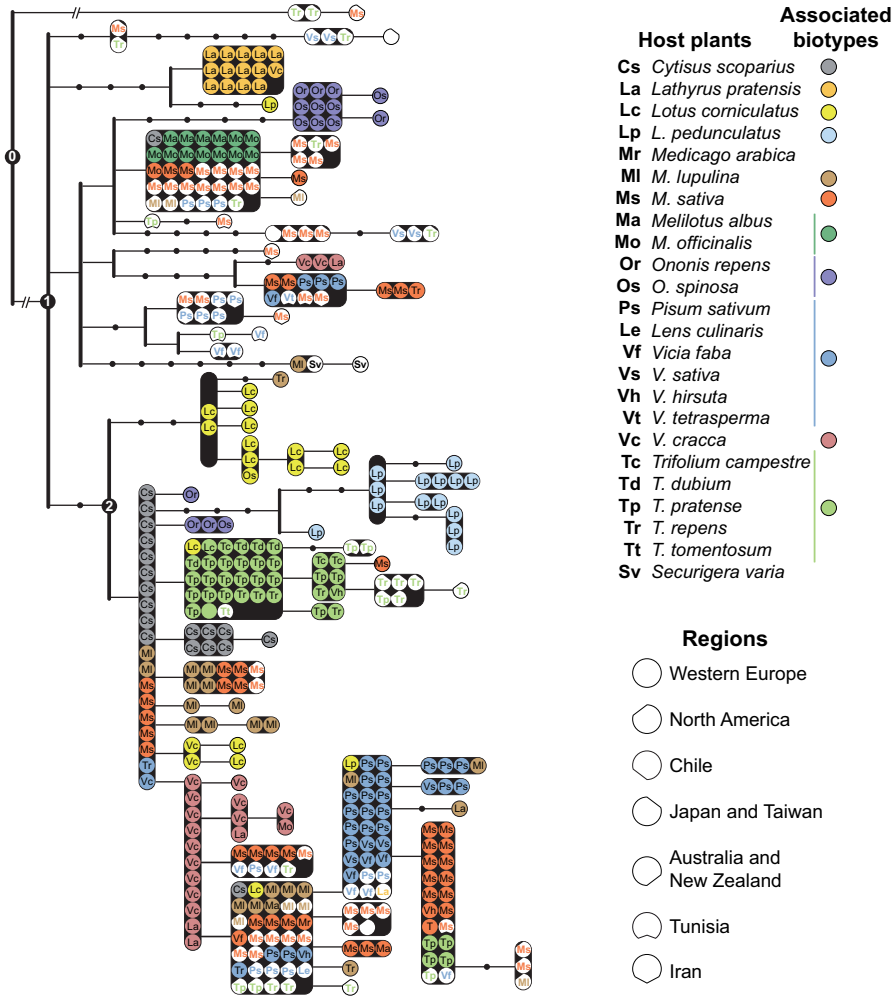

Fig. 1. Phylogeny of pea aphid matrilines from different plants, biotypes, and regions, based on *Buchnera* sequences. The parsimony network connects 365 pea aphids sequenced for 2 *Buchnera* regions. Each individual is represented by a disc-like symbol. The disc shape indicates general provenance, and 2 letters encode the host plant. Individuals previously assigned to specialized biotypes based on microsatellite genotypes (22, 28) are shown in different colors. For other individuals, source biotypes are suggested by the colors of their host plant codes. Individuals within a box share the same haplotype. Each line segment represents 1 mutational step separating haplotypes.

divergence times for 3 main clades. Divergence times were estimated by sampling genealogies generated and weighted under the Bayesian framework implemented in the BEAST program (29). Analyses placed the root in the above tree on the branch leading to the 3 Japanese samples (upper branch in Fig. 1). This rooting was confirmed with sequences of coding regions of 3 *Buchnera* genes using an outgroup, *Buchnera* of the related aphid *Acyrthosiphon kondoi*, and representative sequences for the divergent Japanese lineage and other lineages within the main cluster (Fig. S1).

Based on this rooting, the maximal age of the most basal divergence (clade 0) occurred ≈31,000 years ago [95% highest probability density (HPD), 18,200–46,520 years]. A more recent clade grouping most European biotypes (clade 2) originated ≈6,200 years ago (95% HPD, 3,683–9,492 years). Because the original calibration point was a maximum age (27), more recent divergences cannot be ruled out.

Variation within mitochondrial sequences is even lower than for *Buchnera* sequences of the same matrilines for both the 12S rRNA and its flanking intergenic region (Fig. S2). For the few changes observed, mitochondrial sequences exhibit good congruence with *Buchnera*-Ap phylogenies (Fig. S2), consistent with other evidence supporting strictly maternal inheritance of *Buchnera* and mitochondrial genomes in aphid lineages (24–26, 30).

Similarly, sequence variation among pea aphid biotypes appears to be very low for nuclear genomes, even for noncoding regions (22), preventing any informative comparison with maternal phylogenies. We thus used 11 highly polymorphic nuclear microsatellite loci, previously analyzed for European biotypes (22), to assess whether data from nuclear markers were consistent with the recent diversification implied by *Buchnera* sequences. Specifically, we estimated the average number of mutations per locus since the start of biotype divergence. To achieve this, we fit a model of speciation, the Isolation with Migration (IM) model (31), to our microsatellite data set, using coalescence simulations implemented in the IMa program (32). We did not investigate all 55 possible divergence events between any 2 of the 11 biotypes, but rather selected pairs of biotypes with different levels of genetic distance, host races and cryptic species. Within the limitations of the IM model, which can simulate the

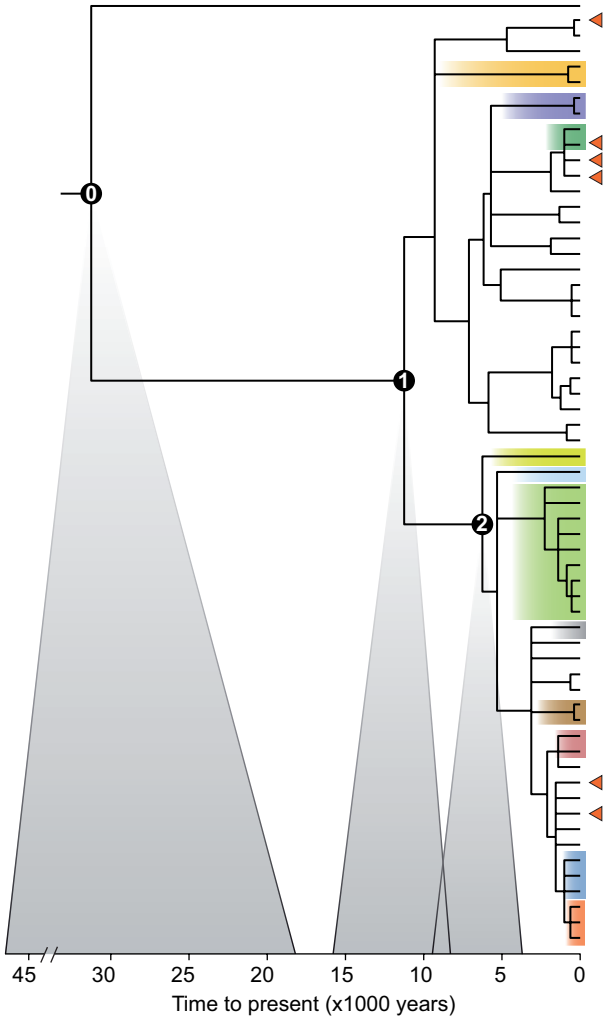

Fig. 2. Phylogeny of major pea aphid matrilines, with branch lengths rescaled to reflect divergence times estimated in the BEAST program. To help visualize the correspondence between phylogenies, major nodes are numbered, and lineages that are mostly associated with particular biotypes are shown over colors corresponding to the colors in Fig. 1. Divergence times were calibrated based on prior dating of node 1, which represents the most recent common ancestor of 7 aphid samples for which *Buchnera* genomes were completely sequenced (27), indicated by red arrowheads. Projections on the time scale represent 95% highest probability densities for the dates of 3 ancestors, as estimated by BEAST.

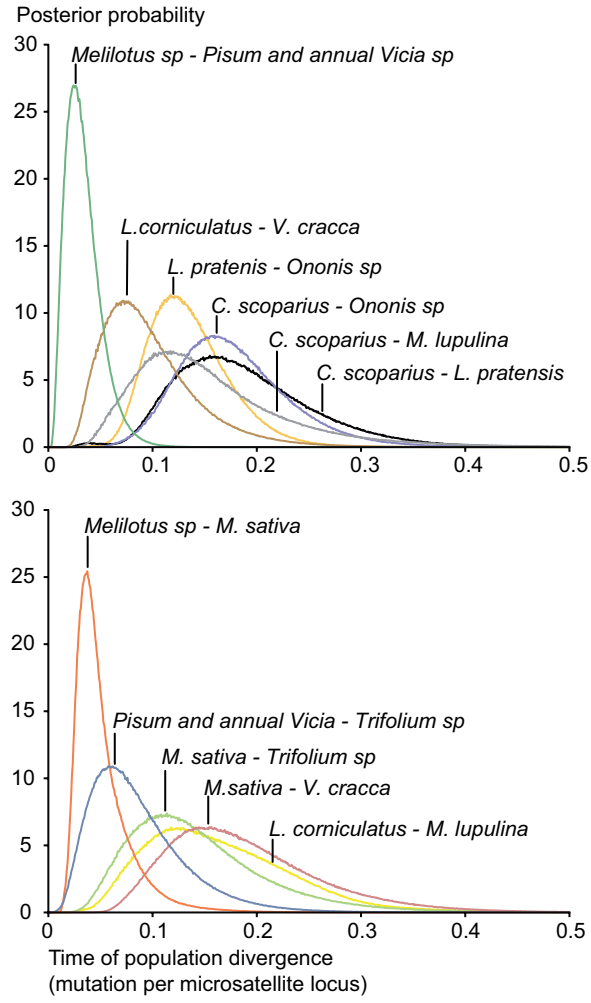

Fig. 3. Estimated divergence times, scaled by the mutation rate of microsatellites, of several pea aphid biotypes named after their typical host plants. Posterior probabilities were estimated by coalescent simulations implemented in the IMa program, fitting a model of population divergence to nuclear microsatellite data obtained previously (22).

divergence with gene flow only between 2 populations in complete isolation from any other population, our analyses consistently produced sharp estimates of divergence between biotypes in the range of 0.15 mutation per microsatellite or less (Fig. 3). Congruence with the diversification beginning some 16,000 years ago would imply a mutation rate on the order of 10^{-5} per locus per year, roughly 10^{-6} per locus per generation of pea aphids, corresponding to laboratory measures of microsatellite mutation rates in other insects (33). Mutation rates of microsatellites in aphids have not yet been estimated (34).

Discussion

Matriline Phylogeny and Recent Biotype Diversification in Pea Aphids. In the matriline phylogeny (Fig. 1), host-specialized biotypes are associated with matrilines that have separated at different times. Because gene divergence generally precedes population divergence (35), this phylogenetic pattern is most readily explained by successive divergence events occurring during the last 16,000 years, the upper bound of the age of the main maternal cluster (Fig. 2). For our findings to be consistent with the hypothesis that the biotypes diverged earlier, we would have to hypothesize a complex history involving successive phases of cytoplasmic introgressions. The good association observed between biotypes and *Buchnera* haplotypes would require periods of restricted maternal gene flow following these waves of introgression. Specifically, the scenario of more ancient biotypes would entail a first wave of introgression spanning most populations 16,000–8,000 years ago to produce the main cluster (clade 1), then a second wave of introgression encompassing all European biotypes except those on *Lathyrus*, *Ononis*, and *Melilotus* ≈6,200 years ago (clade 2, Fig. 2). Although such a scenario involving 2 waves of introgression spanning different sets of biotypes cannot be entirely excluded, recent divergence of biotypes is a more parsimonious explanation for our observations. Furthermore, nuclear microsatellite markers are supportive of very recent divergence (Fig. 3).

The distinct lineage sampled in Japan represents an earlier divergence, occurring 18,000–46,000 years ago (clade 0, Fig. 2). If this lineage simply belonged to a biotype on *Trifolium* or *Medicago sativa*, the plants on which it was sampled on 3 occasions (Fig. 1), then the maintenance of this unique, deeply diverging lineage would not be expected, because ancestral lineages are eliminated through time in the coalescent process. This lineage may possibly represent accidental occurrences on *Trifolium* and *Medicago* of a more anciently diverged biotype or species related to the pea aphid complex.

Recent movements of aphid populations with legume crops introduced to many regions worldwide and our incomplete geographical sampling in Asia prevent us from localizing the region(s) of diversification of pea aphid biotypes with confidence. However, the matriline phylogeny based on *Buchnera* sequences (Fig. 1) provides information on whether pea aphid populations feeding on different crops in colonized countries descend from multiple introductions of already specialized biotypes, as has been reported for Chilean populations (28). For North America, 2 extensively studied races on alfalfa (*M. sativa*) and clover (*Trifolium* sp) (15, 17, 36, 37) are not well represented in our study, but the common ancestry of most of the matrilines collected on clover (Fig. 1) can be parsimoniously explained by the introduction of a clover-specialized biotype to this continent. The same inference could be drawn for Japanese populations on *Trifolium*, which share close maternal ancestry with European ones. This point is not resolved for the alfalfa biotype, because it does not form a distinct matriline.

Diversification Mode and Diversification Rate in Pea Aphids. The origin of the 11 reported pea aphid races and species (22) within the last 16,000 years implies the rapid evolution of reproductive isolation and a minimum of 1 divergence event per lineage every 6,700 years [16,000/ln (11); see ref. 38]. The actual diversification rate in the pea aphid complex is likely higher, because we surveyed only a fraction of its known host range (39) and used conservative calibrations of divergence time (27). The estimated diversification rate is 30-fold higher than the fastest yet reported for arthropods, that for the Hawaiian cricket, *Laupala* (40). It is on par with the fastest speciation rates measured, the radiation of cichlid fishes in African great lakes (38, 41).

Such adaptive radiation cannot reflect codiversification between pea aphids and their host plants, because the latter diversified between 15 and 30 million years ago (42, 43). Instead, the diversification of pea aphids likely involves the acquisition of novel host plants and divergence from the parental lineages remaining on ancestral plants—that is, host shifts (8, 44–46). Ancestral host plants may be inferred by a higher polymorphism of associated matrilines and their paraphyly with respect to those associated with derived hosts (47, 48). In this regard, the phylogeny (Fig. 1) does not show a clear pattern. In addition, because some host-specific biotypes may be unsampled, identifying ancestral and derived host plants of pea aphids without ambiguity would require further studies.

What Caused the Sudden Diversification of Pea Aphid Biotypes? Adaptive radiation is thought to result from a sudden access to underexploited resources and habitats (1, 3). Pea aphid host plants have been present in the general region of western Asia, northern Africa, and Mediterranean Europe for millions of years (42), but it is likely that the availability of these potential resources has increased in the past 10,000 years. The end of the last glacial period resulted in a series of abrupt climate changes 15,000–6,000 years ago across Europe, western Asia, and northern Africa, which had major consequences for the relative abundances of plant species, as indicated by pollen records from Europe (49). Because *Fabaceae* are insect-pollinated plants producing little pollen, few direct data are available regarding their changes in distribution and abundance; however, it is likely that some legumes, among many other plant species, became more abundant or widespread as a result of these climatic transitions.

More direct evidence for an increased range and abundance of several pea aphid hosts is available from archeological data for domesticated or weedy species. *Lens culinaris* (lentil), *Pisum sativum* (pea), and *Vicia* species (broad bean and bitter vetch) were among the original crops brought into domestication in western Asia by the first farmers, starting around 11,500 years ago (50, 51). Other host plants, including species of *Trifolium* (clover) and *Medicago* (alfalfa), were associated with human agriculture as weeds and spread along with the crop species (50). Most of these species spread throughout Europe 8,200–7,000 years ago, where they were cultivated in persistent stands (50, 52, 53). Pollen records indicate the invasion of (wind-pollinated) cereal crops known to be grown with the legume crops or weeds, and also reflect the widespread forest-clearing associated with early agriculture in Europe (54).

The correspondence between the timing of matriline divergences and anthropogenic changes in host plant distributions and abundance presents us with the possibility that Neolithic agriculture contributed to the diversification of the pea aphid complex. Human intervention in host race formation is evidenced by historical observations of host shifts to exotic species (10, 46, 55), but a direct influence of Neolithic agriculture on the diversification of crop parasites (56) awaits validation. As we have shown here, this uncertainty may be addressed with the development of molecular markers calibrated for recent divergence and genomic resources, allowing more precision in estimating the dates of species' origins.

Materials and Methods

Study Material. The study material comprised DNA samples from parthenogenetic female pea aphids, the origins of which are detailed in Table S1. Each sample consisted of a single individual from a separate collection.

Sequences. Sequences were obtained with standard techniques for the full set of samples for 2 regions of the *Buchnera* genome, corresponding to positions 20408–21459 (*groEL-efp*) and 30176–31218 (*cof-metE*) in a published genome for *Buchnera*-Ap (NC_002528; ref. 57). We obtained a third region (positions 578577–579647; *prfC-yhgI*) for 175 samples, the set used for estimating dates of divergence. Overall, 83% of the sites analyzed were within intergenic regions. Sequencing of the mitochondrial region corresponded to the 3′ end of the 12S rRNA plus downstream spacer (59 samples, positions 14481–14930 of FJ411411). Because spacer regions between *Buchnera*-Ap and the outgroup *Buchnera*-Ak were too divergent to allow reliable alignment, we used coding sequences of *Buchnera* genes *dapB*, *dxr*, and *yidC* for the purpose of rooting. Table S2 lists the markers and primers used.

Tree Construction. *Buchnera* sequences were concatenated for each individual and aligned using BioEDIT or MacClade (58). In cases of partial missing data for particular samples, we used the assumption that the missing nucleotides matched the most closely related haplotype. We considered the insertions or deletions of several adjacent nucleotides in a sequence to be the result of a single mutational event.

Indels associated with homopolymers or repetitive regions are known to be recurrent mutations in *Buchnera*-Ap (27, 59), and in cases of conflict with other sites, we assumed that such indels occurred multiple times. In a single instance, we also favored transitions over transversions at the same site. All other polymorphic sites were unambiguous, resulting in a single most parsimonious tree, built by heuristic search in PAUP* 4.0b10 (60) under default settings. A parsimony network of all individuals sequenced at *cof-metE* and *groEL-efp* regions (1,596 aligned nucleotides) was then built in TCS 2.1 (61) (Fig. 1). For the few cases of ambiguous loops in the network, we favored paths that were congruent with the topology of the tree based on the 3 loci.

Estimation of Ancestral Dates. The Bayesian Markov chain Monte Carlo (MCMC) method to estimate divergence times, as implemented in BEAST 1.4.4 (29, 62), was used with the alignment of unique haplotypes corresponding to the 3 *Buchnera* loci. Loci were assumed to evolve under the HKY model of substitution, with estimated base frequencies and no site heterogeneity. Because of the low sequence divergence ($<$0.6%), varying models of substitution would have essentially no effect on the results.

We calibrated the phylogeny using the estimated divergence time among 7 sequenced *Buchnera* genomes, specified as a uniform distribution from 8,340 to 15,790 years (27). This divergence time was derived from 2 independent calibration points based on descendants of 2 female ancestors introduced to North America on or after 1870, based on historical records. The 2 introduced clusters gave nearly identical divergence rates, which also were consistent with the divergence of *Buchnera* in laboratory lines derived from a single female (27). Because the ancestor for each cluster could have occurred after the first historical record of pea aphids, the divergence dates should be considered maxima. The primary assumption underlying the calibration is that very recent divergences (during the last 135 years) have the same base substitution rate as the deeper divergences (over thousands of years). Because calculations based on silent sites give the same date estimates (27), we could rule out the possibility that very recent evolution is unrepresentative due to recent acquisition of deleterious mutations that are eliminated after longer periods.

As a demographic model, we specified the Bayesian Skyline Plot (63), which minimizes assumptions on the demographic history of populations. We ran analyses for 10^7 generations, sampling every 1,000th iteration after an initial burn-in of 10^6 steps. We used Tracer v1.4 (64) to check the performance of the MCMC and to depict the posterior probability distributions of parameters. Preliminary analyses allowing for a relaxed molecular clock (62) with uncorrelated, branch-specific rates following a lognormal distribution did not reject the hypothesis of constant molecular evolution of the sequences. Thus, we assumed a strict clock, and combined the results of the last 9,000 sampled trees of 2 runs using LogCombiner and TreeAnnotator. We used the parsimony tree built in PAUP* (see above) as the target topology and rescaled branches to reflect the posterior mean node heights.

Estimations of Population Divergence Time with Coalescence Simulations. We used microsatellite data at 11 loci [*AlA12M, AlB04M, AlB07M, AlB08M, AlB12M, ApH08M, ApH10M* (65), *Ap-03* (66), *S23, S30* (67), and *Sm11* (68)] previously analyzed in aphids collected in eastern France on several host plants and genetically assigned to 1 of 10 host-specialized biotypes (22) (Table S1). We used aphids from a single location so as to minimize the effects of genetic structuring within populations, which deviates from IMa's assumptions. A single biotype, on *Lotus pedunculatus*, was not sampled in this region and thus was not included in the analyses. Microsatellites were assumed to evolve via stepwise mutations. Suitable priors for model parameters (i.e., population sizes, migration rates, and divergence time) were chosen after short IMa runs, as recommended in the program manual. To ensure proper mixing, primary runs used 10 coupled MCMC with a geometric heating scheme of parameters $g1 = 0.9$ and $g2 = 0.98$. We discarded the first 10^6 MCMC steps and ended runs after 10 days. At this time, MCMC had reached between $4*10^7$ and $5*10^7$ steps, parameters had effective sample sizes $>$50, and their evolution over the course of simulations showed no visible trend in the trendline plots.

ACKNOWLEDGMENTS. We thank the following people who provided aphid samples or helped with their collection: Saghar Aleosfoor, Paul Baumann, Emilie Bilodeau, Gaelen Burke, Marina Caillaud, Steve Clement, Conrad Cloutier, Angela Douglas, Sanford Eigenbrode, Owain Edwards, Christian Figueroa, Robert Foottit, Takema Fukatsu, Nicole Gerardo, Hanene Harbaoui, Anthony Ives, Blas Lavandero, Weinong Lu, Hanem Makni, Susan Masta, Roberto Nespolo, Kerry Oliver, Yannick Outreman, Servane Penvern, Manuel Plantegenest, Glen Powell, Claudio Ramirez, Jacob Russell, Beatriz Sabater-Muñoz, Andrea Silva Baez. Tsutomu Tsuchida, Wolfgang Weisser, Alex Wilson, and Steve Wratten. We also thank Lucie Mieuzet, Morgane Perrenou, and Laurel Johnstone for their support with the molecular analyzes, and Manuel Plantegenest and Joel Wertheim for their comments on a draft of the manuscript. Funding was provided by U.S. National Science Foundation Grant 0723472 (to N.A.M.) and by the French ''Ecologie pour la Gestion des Ecosystèmes et de leurs Ressources'' (ECOGER) program (to J.-C.S.).

1. Schluter D (2000) *The Ecology of Adaptive Radiation* (Oxford Univ Press, Oxford).
2. Simpson GG (1953) *The Major Features of Evolution* (Columbia Univ Press, New York).
3. Gavrilets S, Losos JB (2009) Adaptive radiation: Contrasting theory with data. *Science* 323:732–737.
4. Rundle HD, Nosil P (2005) Ecological speciation. *Ecol Lett* 8:336–352.
5. Schluter D (2001) Ecology and the origin of species. *Trends Ecol Evol* 16:372–380.
6. Terai Y, et al. (2006) Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. *PLoS Biol* 4:2244–2251.
7. Gavrilets S, Vose A (2005) Dynamic patterns of adaptive radiation. *Proc Natl Acad Sci USA* 102:18040–18045.
8. Berlocher SH, Feder JL (2002) Sympatric speciation in phytophagous insects: Moving beyond controversy? *Annu Rev Entomol* 47:773–815.
9. Drès M, Mallet J (2002) Host races in plant-feeding insects and their importance in sympatric speciation. *Philos Trans R Soc Lond Ser B* 357:471–492.
10. Carroll SP, Boyd C (1992) Host race radiation in the soapberry bug: Natural history with the history. *Evolution* 46:1052–1069.
11. Feder JL, et al. (2003) Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. *Proc Natl Acad Sci USA* 100:10314–10319.
12. Caillaud MC, Via S (2000) Specialized feeding behavior influences both ecological specialization and assortative mating in sympatric host races of pea aphids. *Am Nat* 156:606–621.
13. Feder JL, et al. (1994) Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly. *Proc Natl Acad Sci USA* 91:7990–7994.
14. Linn CE, et al. (2004) Postzygotic isolating factor in sympatric speciation in *Rhagoletis* flies: Reduced response of hybrids to parental host-fruit odors. *Proc Natl Acad Sci USA* 101:17753–17758.
15. Via S, Bouck AC, Skillman S (2000) Reproductive isolation between divergent races of pea aphids on two hosts, II: Selection against migrants and hybrids in the parental environments. *Evolution* 54:1626–1637.
16. Müller FP (1962) Biotypes and subspecies of the ''pea louse'' *Acyrthosiphon pisum* (Harris) (Translated from German). *Z Pflanzenkrankheite* 69:129–136.
17. Via S (1991) The genetic structure of host plant adaptation in a spatial patchwork: Demographic variability among reciprocally transplanted pea aphid clones. *Evolution* 45:827–852.
18. Ferrari J, Godfray HCJ, Faulconbridge AS, Prior K, Via S (2006) Population differentiation and genetic variation in host choice among pea aphids from eight host plant genera. *Evolution* 60:1574–1584.
19. Ferrari J, Via S, Godfray HCJ (2008) Population differentiation and genetic variation in performance on eight hosts in the pea aphid complex. *Evolution* 62:2508–2524.
20. Frantz A, Plantegenest M, Mieuzet L, Simon J-C (2006) Ecological specialization correlates with genotypic differentiation in sympatric host populations of the pea aphid. *J Evol Biol* 19:392–401.
21. Simon J-C, et al. (2003) Host-based divergence in populations of the pea aphid: Insights from nuclear markers and the prevalence of facultative symbionts. *Proc R Soc Lond Ser B* 270:1703–1712.

22. Peccoud J, Ollivier A, Plantegenest M, Simon J-C (2009) A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proc Natl Acad Sci USA* 16:7495–7500.
23. Birkle LM, Douglas AE (1999) Low genetic diversity among pea aphid (*Acyrthosiphon pisum*) biotypes of different plant affiliation. *Heredity* 82:605–612.
24. Funk DJ, Helbling L, Wernegreen JJ, Moran NA (2000) Intraspecific phylogenetic congruence among multiple symbiont genomes. *Proc R Soc Lond Ser B* 267:2517–2521.
25. Jousselin E, Desdevises Y, Coeur d'acier A (2009) Fine-scale cospeciation between *Brachycaudus* and *Buchnera aphidicola*: Bacterial genome helps define species and evolutionary relationships in aphids. *Proc R Soc Lond Ser B* 276:187–196.
26. Lozier JD, Roderick GK, Mills NJ (2007) Genetic evidence from mitochondrial, nuclear, and endosymbiont markers for the evolution of host plant associated species in the aphid genus *Hyalopterus* (Hemiptera: Aphididae). *Evolution* 61:1353–1367.
27. Moran NA, McLaughlin HJ, Sorek R (2009) The dynamics and timescale of ongoing genomic erosion in symbiotic bacteria. *Science* 323:379–382.
28. Peccoud J, et al. (2008) Host range expansion of an introduced insect pest through multiple colonizations of specialized clones. *Mol Ecol* 17:4608–4618.
29. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
30. Abbot P, Withgott JH (2004) Phylogenetic and molecular evidence for allochronic speciation in gall-forming aphids (*Pemphigus*). *Evolution* 58:539–553.
31. Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis. Genetics* 167:747–760.
32. Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc Natl Acad Sci USA* 104:2785–2790.
33. Vazquez JF, Perez T, Albornoz J, Dominguez A (2000) Estimation of microsatellite mutation rates in *Drosophila melanogaster. Genet Res* 76:323–326.
34. Wilson ACC, Sunnucks P, Hales DF (2002) Heritable genetic variation and potential for adaptive evolution in asexual aphids (Aphidoidea). *Biol J Linn Soc* 79:115–135.
35. Avise JC (2000) *Phylogeography: The History and Formation of Species* (Harvard Univ Press, Cambridge, MA), p 447.
36. Hawthorne DJ, Via S (2001) Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature* 412:904–907.
37. Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol Ecol* 17:4334–4345.
38. Coyne JA, Orr HA (2004) *Speciation* (Sinauer, Sunderland, MA), p 545.
39. Eastop VF (1971) Keys for the identification of *Acyrthosiphon* (Hemiptera: Aphididae). *Bull Br Mus (Nat Hist) Entomol* 26:1–115.
40. Mendelson TC, Shaw KL (2005) Sexual behaviour: Rapid speciation in an arthropod. *Nature* 433:375–376.
41. Verheyen E, Salzburger W, Snoeks J, Meyer A (2003) Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. *Science* 300:325–329.
42. Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of *Leguminosae* implicates a rapid diversification of lineages during the Tertiary. *Syst Biol* 54:575–594.
43. Maureira-Butler IJ, Peeil BE, Muangprom A, Osborn TC, Doyle JJ (2008) The reticulate history of *Medicago* (Fabaceae). *Syst Biol* 57:466–482.
44. Agosta SJ (2006) On ecological fitting, plant–insect associations, herbivore host shifts, and host plant selection. *Oikos* 114:556–565.
45. Bush GL (1969) Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera: Tephritidae). *Evolution* 23:237–251.
46. Groman P (2000) Rapid evolution and specialization following host colonization in a yucca moth. *J Evol Biol* 13:223–236.
47. Brown JM, Abrahamson WG, Way PA (1996) Mitochondrial DNA phylogeography of host races of the goldenrod ball gallmaker, *Eurosta solidaginis* (Diptera: Tephritidae) *Evolution* 50:777–786.
48. Ohshima I (2008) Host race formation in the leaf-mining moth, *Acrocercops transecta* (Lepidoptera: Gracillariidae). *Biol J Linn Soc* 93:135–145.
49. Birks HH, Ammann B (2000) Two terrestrial records of rapid climatic change during the glacial–Holocene transition (14,000–9,000 calendar years BP) from Europe. *Proc Natl Acad Sci USA* 97:1390–1394.
50. Colledge S, Conolly J (2007) *The Origins and Spread of Domestic Plants in Southwest Asia and Europe* (West Coast Press, Walnut Creek, CA).
51. Zohary D, Hopf M (2000) *Domestication of Plants in the Old World* (Clarendon, Oxford), 3rd Ed.
52. Colledge S, Conolly J, Shennan S (2005) The evolution of neolithic farming from SW Asian origins to NW European limits. *Eur J Archaeol* 8:137–156.
53. Zeder MA (2008) Domestication and early agriculture in the Mediterranean basin: Origins, diffusion, and impact. *Proc Natl Acad Sci USA* 105:11597–11604.
54. Wehrli M, Tinner W, Ammann B (2007) 16 000 years of vegetation and settlement history from Egelsee (Menzingen, central Switzerland). *Holocene* 17:747–761.
55. Walsh B (1864) On phytophagic varieties and phytophagic species. *Proc Entomol Soc Phil* 3:403–430.
56. Zaffarano PL, McDonald BA, Linde CC (2008) Rapid speciation following recent host shifts in the plant pathogenic fungus *Rhynchosporium. Evolution* 62:1418–1436.
57. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H (2000) Genome sequence of the endocellular bacterial symbiont of aphids, *Buchnera* sp APS. *Nature* 407:81–86.
58. Maddison DR, Maddison WP (2000) *MacClade 4: Analysis of Phylogeny and Character Evolution* (Sinauer Associates, Sunderland, MA).
59. Dunbar HE, Wilson ACC, Ferguson NR, Moran NA (2007) Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS Biol* 5:e96.
60. Swofford DL (2000) *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods*) (Sinauer Associates, Sunderland, MA).
61. Clement M, Posada D, Crandall KA (2000) TCS: A computer program to estimate gene genealogies. *Mol Ecol* 9:1657–1659.
62. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4:699–710.
63. Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22:1185–1192.
64. Rambaut A, Drummond AJ (2007) Tracer v1.4. Available from http://beast.bio.ed.ac.uk/Tracer.
65. Caillaud MC, et al. (2004) Microsatellite DNA markers for the pea aphid, *Acyrthosiphon pisum. Mol Ecol Notes* 4:446–448.
66. Kurokawa T, Yao I, Akimoto SI, Hasegawa E (2004) Isolation of six microsatellite markers from the pea aphid, *Acyrthosiphon pisum* (Homoptera: Aphididae) *Mol Ecol Notes* 4:523–524.
67. Wilson ACC, et al. (2004) Cross-species amplification of microsatellite loci in aphids: Assessment and application. *Mol Ecol Notes* 4:104–109.
68. Simon JC, et al. (1999) Reproductive mode and population genetic structure of the cereal aphid, *Sitobion avenae*, studied using phenotypic and microsatellite markers. *Mol Ecol* 8:531–545.